Annual Progress Report

Extreme-Scale Distribution-Based Data Analysis (EDDA)

The Ohio State University

Award #: SN10015 DE-SC00124954

Report period: September 2014 - June 2015

Principle Investigator: Han-Wei Shen

Major Research Activities and Findings:

The goal of this project is to develop a data analysis and visualization framework for efficient and compact representations of extreme-scale scientific data sets. Compared to the traditional data reduction approaches, distribution-based methods allow one to more accurately capture the underlying statistical properties of the raw data, and to precisely model the errors resulted from the process of data transformation and reduction. The statistical properties captured by our framework make it possible to produce visualization of better quality and to model the uncertainty. Certain statistics extracted from the data also allow one to identify salient features for specific applications. In the first year of this project, we have developed several novel algorithms that can achieve:

- Time-varying vector data reduction and better uncertain pathline computation
- Particle tracing in uncertain vector fields based on particle and Kalman filters
- Local histogram computation and query based on bitmap indices
- Statistical feature extraction for time-varying and vortex detection and analysis
- Association analysis of multivariate scientific data sets

Our research produced multiple high quality research papers submitted and accepted to high quality visualization conferences, including three IEEE Pacific Visualization 2015 papers, one EuroVis 2015 paper, and seven submissions to IEEE SciVis 2015. Out of the seven submissions to IEEE SciVis 2015, three have been notified being conditionally accepted as of 6/7/2015. Besides the publications, the PI Han-Wei Shen organized a featured SIAM mini-symposium on big data analytics in SIAM CSE 2015, participated a panel on “what is data science” in SIAM CSE 2015 led by Prof. Chris Johnson at University of Utah.

Below we highlight some representative results of our research.

1. Uncertainty Modeling and Error Recution for Pathline Computation (Pacific Vis 2015)

We devised a novel error modeling and reduction algorithm to model temporal interpolation errors and improve the data accuracy for temporally down-sampled data. We show that it is possible to compute polynomial regression and measure the interpolation errors incrementally with one sequential scan of the time-varying flow field. We also show empirically that when the data sequence is fitted with least-squares regression, the errors can be approximated with a Gaussian distribution. With the end positions of particle traces stored, we show that our error modeling scheme can better estimate the intermediate particle trajectories between the stored time steps based on a maximum likelihood method that utilizes forward and backward particle traces.

2. Uncertainty-Driven Vortex Analysis (Pacific Vis 2015)

In this work, we developed a consensus-based uncertainty model and introduce spatial proximity to enhance vortex detection results obtained using point-based methods. We use four existing local vortex detectors and convert their outputs into fuzzy possibility values using a sigmoid-based soft-thresholding approach. We apply a majority voting scheme that enables us to identify candidate vortex regions with a higher degree of confidence. Then, we introduce spatial proximity- based analysis to discern the final vortical regions. Thus, by using spatial proximity coupled with fuzzy inputs, we propose a novel uncertainty analysis approach for vortex detection. We use expert’s input to better estimate the system parameters and results from two real-world data sets demonstrate the efficacy of our method.

3. Efficient Local Historgram Search via Bitmap Indexing (EuroVis 2015)

We developed a novel algorithm to accelerate local histogram search by leveraging bitmap indexing. Our method avoids exhaustive searching of all voxels in the spatial domain by examining only the voxels whose values fall into the value range of user-defined local features and their neighborhood. Based on the idea that the value range of local features is in general much smaller than the dynamic range of the entire data set, we propose a local voting scheme to construct the local histograms so that only a small number of voxels need to be examined. Experimental results showed that our method can reduce the computational workload by two orders of magnitudes compared to the conventional approaches while keeping the memory overhead low. An interactive interface was developed to assist users to define target features as local histograms and identify the location of these features in the data

4. Visualization and Analysis of Rotating Stall for Transonic Jet Engine Simulations (Conditionally accepted to IEEE SciVis 2015)

Identification of early signs of rotating stall is essential for the study of turbine engine stability. With recent advancements of high performance computing, high-resolution unsteady flow fields allow in depth exploration of rotating stall and its possible causes. Performing stall analysis, however, involves significant effort to process large amounts of simulation data, especially when investigating abnormalities across many time steps. In order to assist scientists during the exploration process, we present a visual analytics framework to identify suspected spatiotemporal regions through a comparative visualization so that scientists are able to focus on relevant data in more detail. To achieve this, we propose efficient stall analysis algorithms derived from domain knowledge and convey the analysis results through juxtaposed interactive plots. Using our integrated visualization system, scientists can visually investigate the detected regions for potential stall initiation and further explore these regions to enhance the understanding of this phenomenon. Positive feedback from scientists demonstrates the efficacy of our system in analyzing rotating stall.

5. Distribution Driven Extraction and Tracking of Features for Time-Varying Data Analysis (Conditionally accepted to IEEE SciVis 2015)

Effective analysis of features in time-varying data is essential in numerous scientific applications. Feature extraction and tracking are two important tasks scientists rely upon to get insights about the dynamic nature of the large scale time-varying data. However, often the complexity of the scientific phenomena only allows scientists to vaguely define their feature of interest. Furthermore, such features can have varying motion patterns and dynamic evolution over time. As a result, automatic extraction and tracking of features becomes a non-trivial task. In this work, we investigate these issues and propose a distribution driven approach that allows us to construct novel algorithms for reliable feature extraction and tracking with high confidence in the absence of accurate feature definition. We exploit two key properties of an object, motion and similarity to the target feature, and fuse the information gained from these two key properties to generate a robust feature-aware classification field at every time step. Tracking of features is done using such classified fields which enhance the accuracy and robustness of the proposed algorithm. The efficacy of our method is demonstrated by successfully applying it on several scientific data sets containing a wide range of dynamic time-varying features.

6. Association Analysis for Visual Exploration of Multivariate Scientific Data Sets (Conditionally accepted to IEEE SciVis 2015)

The heterogeneity and complexity of multivariate characteristics poses a unique challenge to visual exploration of multivariate scientific data sets, as it requires investigating the usually hidden associations between different variables and specific scalar values to understand the data's multi-faceted properties. In this work , we design a novel association analysis method that guides visual exploration of scalar-level associations in the multivariate context. We model the directional interactions between scalars of different variables as information flows based on association rules. We introduce the concepts of informativeness and uniqueness to describe how information flows between scalars of different variables and how they are associated with each other in the multivariate domain. Based on scalar-level associations represented by a probabilistic association graph, we propose the Multi-Scalar Informativeness-Uniqueness (MSIU) algorithm to evaluate the informativeness and uniqueness of scalars. We present an exploration framework with multiple interactive views to explore the scalars of interest with confident associations in the multivariate spatial domain, and provide guidelines for visual exploration using our framework. We demonstrate the effectiveness and usefulness of our approach through case studies using three representative multivariate scientific data sets.

7. Probabilistic Streamline Computation using Particle Filtering (Submitted to IEEE SciVis 2015)

Streamline-based techniques play an important role in visualizing and analyzing uncertain steady vector fields. Generating accurate streamlines in uncertain vector fields is a challenging problem due to the rapid propagation of errors in global regions. In this work, we developed a novel probabilistic method for streamline computation on distribution-based steady vector fields using a particle filtering framework. In our framework, a streamline is modeled as a state space model that captures the spatial coherence of integration steps. Distributions in adjacent data blocks are used to provide the conditional prior density and the likelihood function. To approximate the posterior distribution for all the possible particle traces originating from a given seed position, a set of weighted samples are iteratively updated from which streamlines with higher likelihood can be derived. We qualitatively and quantitatively compare our method with alternative methods on different types of flow field data sets. Our method can generate possible streamlines with higher certainty and hence more accurate flow traces.

Publications

1. Lu, Kewei, Shen, Han-Wei, Peterka, Tom. "Scalable Computation of Stream Surfaces on Large Scale Vector Fields" In ACM Supercomputing 2014. New Orleans, November 2014.

2. Tong, Xin, Chen, Chun-Ming, Shen, Han-Wei, Wang, Pak. "Interactive Streamline Exploration and Manipulation Using Deformation" In IEEE Pacific Visualization 2015. Hangzhou, April 2015

3. Liu, Xiaotong, Shen, Han-Wei. "The Effects of Representation and Juxtaposition on Graphical Perception of Matrix Visualization" In ACM CHI. Seoul, Korea, ACM, May 2015

4. Chen, Chun-Ming, Biswas, Ayan, Shen, Han-Wei. "Uncertainty Modeling and Error Reduction for Pathline Computation in Time-varying Flow Fields" In IEEE Pacific Visualization 2015. Hangzhou, April 2015

5. Biswas, Ayan, Thompson, David, He, Wenbin, Deng, Qi, Chen, Chun-Ming, Shen, Han-Wei,

Machiraju, Raghu, Rangarjan, Anand. "An Uncertainty-Driven Approach to Vortex Analysis Using Oracle Consensus and Spatial Proximity" In IEEE Pacific Visualization 2015. Hangzhou, April 2015

6. Wei, Tzu-Hsuan, Chen, Chun-Ming, Biswas, Ayan, "Efficient Local Histogram Searching via Bitmap Indexing", in EuroVis 2015, Italy, May 2015